104-02717/

Bank of Canada Working Paper 2004-5

February 2004

Structural Change and Forecasting Long-Run Energy Prices

by

Jean-Thomas Bernard,[1] Lynda Khalaf,[1] and Maral Kichian[2]

[1]Département d'économique and GREEN
Université Laval
Québec, Canada G1K 7P4

[2]Research Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
mkichian@bankofcanada.ca

The views expressed in this paper are those of the authors.
No responsibility for them should be attributed to the Bank of Canada.

Contents

Acknowledgements iv
Abstract/Résumé v

1. Introduction 1
2. Models and Test Strategy 2
3. Empirical Results 7
4. Conclusion 12

Bibliography 13
Appendix A: The Maximized Monte Carlo Test 16
Appendix B: Kalman Filtering and the TVP Model 18

Acknowledgements

We would like to thank Jean-Marie Dufour and Eva Ortega for helpful suggestions and comments, and Robert Pindyck for kindly providing the data set.

Abstract

The authors test the statistical significance of Pindyck's (1999) suggested class of econometric equations that model the behaviour of long-run real energy prices. The models postulate mean-reverting prices with continuous and random changes in their level and trend, and are estimated using Kalman filtering. In such contexts, test statistics are typically non-standard and depend on nuisance parameters. The authors use simulation-based procedures to address this issue; namely, a standard Monte Carlo test and a maximized Monte Carlo test. They find statistically significant instabilities for coal and natural gas prices, but not for crude oil prices. Out-of-sample forecasts are calculated to differentiate between significant models.

JEL classification: C22, C52, C53, Q40
Bank classification: Econometric and statistical methods

Résumé

Les auteurs testent la signification statistique de la famille d'équations économétriques que Pindyck (1999) met en avant pour modéliser le comportement des prix réels de l'énergie en longue période. Ces modèles, qu'il estime au moyen du filtre de Kalman, postulent la stationnarité des prix par rapport à la moyenne ainsi que des variations continues et aléatoires de leur niveau et de la pente de leur tendance. Dans ces conditions, les statistiques de test ne sont généralement pas standard et dépendent de paramètres de nuisance. Afin de contourner la difficulté, les auteurs recourent à des procédures de test reposant sur la simulation, en l'occurrence les versions standard et maximisée du test de Monte-Carlo. Ils constatent des instabilités statistiquement significatives dans le cas des cours du charbon et du gaz naturel, mais non dans celui du pétrole brut. Ils font des prévisions en dehors de la période d'estimation dans le but de départager les modèles significatifs.

Classification JEL : C22, C52, C53, Q40
Classification de la Banque : Méthodes économétriques et statistiques

1. Introduction

It has long been known that fluctuations in energy prices have important and lasting effects on the economies of industrialized countries. As a recent example, Hamilton (2003) finds a strongly significant and non-linear relationship between changes in oil prices and GDP growth. Similarly, in small open-economy settings, Amano and van Norden (1995) find long-run links between oil prices and real exchange rates. Enduring price movements in energy commodities can also cause relative price changes among a wide range of products in the economy, ultimately feeding into the rate of inflation for some duration. Thus, future values of oil prices in particular, and energy products in general, are important ingredients of long-run forecasts for various macroeconomic variables.

Forecasting the behaviour of energy prices can be quite challenging. In addition to domestic and international supply and demand conditions, a complete model also needs to take into account market regulations, technological advances, and geopolitical considerations. These non-market-related aspects present the biggest challenges for the forecaster, since they are largely unpredictable. For that reason, Pindyck (1999) suggests that, rather than fully articulated structural equations, it is preferable to adopt simple models – where prices grow in real terms and at a fixed rate – for our long-run forecasting needs.

Despite being simple, these models are flexible, allowing prices to grow from their current level (i.e., prices follow a random-walk process with drift) and/or from a changing trend line (i.e., prices revert to a possibly moving mean). Such differences can be thought of as reflecting differing assumptions regarding resource depletion and technological change.[1] Indeed, using a simple Hotelling model, Pindyck shows that long-run energy prices should revert to an *unobservable* trending long-run marginal cost, with continuous random changes in their level and in the slope of their trend. He proposes a family of econometric models for these prices (discrete versions of multivariate Ornstein-Uhlenbeck processes) that integrate the desired features. Versions of these models are estimated using the Kalman filter, and out-of-sample forecasts are obtained.

The forecast excercises conducted by Pindyck show some mixed results, but overall

[1] For example, an energy type that is produced and sold competitively should have its price reverting to a long-run marginal cost that itself can change over time.

good potential, for the class of equations considered. The estimated models, however, are not tested statistically. Furthermore, the statistical significance of continuous and random changes in the level and trend of energy prices *in general* is not addressed. Yet, whether time-varying-parameter (TVP) approaches to long-run forecasting are successful depends importantly (although not exclusively) on their statistical goodness-of-fit. In other words, it would be interesting to determine whether Pindyck's mixed results can be rationalized statistically.

In this paper, we complement Pindyck's study by conducting tests for continuous and random shifts in real energy prices. We are thus able to select, within the suggested family of models, specifications that are statistically significant for crude oil, coal, and natural gas prices. Our methodology relies on simulation-based exact testing procedures, applicable in situations where standard testing is not valid; that is, when unidentified nuisance parameters are present under the null hypothesis.[2] The tests are also valid when data samples are small. We complete our results by providing a number of out-of-sample forecast statistics.

Our findings indicate significant TVP effects in two of the three energy-price series examined, which supports Pindyck's proposed class of models. Indeed, we find a multiplicity of significant TVP specifications in the natural gas and coal series. To distinguish between them, we contrast the forecasting ability of the TVP models with forecasts obtained from a fixed-coefficient model and from a random-walk-with-drift model.

Section 2 describes the class of proposed models and the testing details. Section 3 documents and discusses the obtained results. Section 4 concludes.

2. Models and Test Strategy

Pindyck (1999) considers a basic Hotelling model for any depletable resource that is produced in a competitive market. With constant marginal costs of extraction, c, and an isoelastic demand function with unitary elasticity, the price level is given by

$$P_t = c + (ce^{rt}/(e^{rcR_0/A} - 1)), \tag{1}$$

[2]These techniques are applied in Khalaf and Kichian (2002).

where R_0 is the level of initial reserves, A is a demand shifter, and r is the interest rate. This implies that the slope of the price trajectory will be given by

$$dP_t/dt = rce^{rt}/(e^{rcR_0/A} - 1), \tag{2}$$

so that changes in demand, extraction costs, and reserves all affect this slope.[3] If, as Pindyck (1999) argues, these factors fluctuate in a continuous and unpredictable manner over time, then long-run energy prices should revert to a trend that itself fluctuates in the same fashion.

A class of models that integrates the above features is a generalized Ornstein-Uhlenbeck process, and Pindyck (1999) proposes its discrete version as a suitable econometric framework for forecasting long-run energy prices. This is given by the AR(1)-based dynamic model:

$$P_t = c_1 + \phi_{1t} + \phi_{2t}t + c_2 P_{t-1} + \epsilon_t, \ t = 1, ..., T, \tag{3}$$

where P_t refers to the log real price of an energy product and the coefficients ϕ_{1t} and ϕ_{2t} follow the stochastic processes

$$\begin{aligned} \phi_{1t} &= c_3\phi_{1,t-1} + v_{1t}, \\ \phi_{2t} &= c_4\phi_{2,t-1} + v_{2t}. \end{aligned}$$

The underlying error terms ϵ_t, v_{1t}, and v_{2t}, $t = 1, ..., T$, are assumed to be independently and identically distributed (i.i.d.) with zero means and variances σ^2_ϵ, $\sigma^2_{v_1}$, and $\sigma^2_{v_2}$, respectively. Following Pindyck's theoretical development and his discussion on unit root tests, we adopt a general mean-reverting framework. That is, we impose $|c_2| < 1$ for all TVP models. The processes for ϕ_{1t} and ϕ_{2t} may or may not be mean reverting. These are the unobservable, continuously evolving parameters that describe long-run marginal costs of the underlying structural model.

Assuming normality for the distributions of ϵ_t, v_{1t}, and v_{2t}, Pindyck proposes that Kalman filtering be applied to obtain paths for the state variables ϕ_{1t} and ϕ_{2t}. This

[3]For example, an increase in A causes this slope to increase, while increases in c or R_0 cause the slope to decrease. In addition, increases in c or A cause the price level to increase, whereas an increase in R_0 leads to a decrease in this level.

means that, starting with initial values for the model parameters and the state variables, each period the filter will calculate new values for the state variables to reflect new information on the observable series. Once the full paths of the state variables are determined, the model can be estimated by maximum likelihood to obtain the optimal parameter estimates. For the model above, details of the Kalman filtering procedure are described in Appendix B.

For the purposes of examining the statistical significance of TVP effects, the null hypothesis of interest is a simple mean-reverting model around a fixed trend line (the trending Ornstein-Uhlenbeck process given by equation (24) in Pindyck 1999), or, in our notation:

$$P_t = c_1 + \phi_1 + \phi_2 t + c_2 P_{t-1} + \epsilon_t, \ t = 1, ..., T. \tag{4}$$

It is clear that the models to be compared statistically are nested at the boundaries of certain parameters; intuitively, the alternative model becomes more and more like the null model as the variances of the time-varying parameters approach zero and the parameters c_3 and c_4 get close to 1. But, since the variance cannot become identically zero, the models are nested at the boundary of these parameters. Formally,

$$\text{model (4)} \sqsubseteq \text{model (3) when } \sigma^2_{v_1} \to 0, \sigma^2_{v_2} \to 0, \text{and } c_3 = c_4 = 1.$$

In the presence of such identification problems, one cannot rely on estimated standard errors, since their use for confidence set and t-test purposes is not justified even asymptotically.[4] Nevertheless, a likelihood-ratio type statistic can be constructed to test the statistical significance of the time variation in the parameters. Specifically, the quasi-likelihood-ratio (QLR) type statistic takes the form

$$QLR(J) = 2[L_{TVP}(J) - L_{FCM}], \tag{5}$$

where $L_{TVP}(J)$ and L_{FCM} are, respectively, the maximum of the log-likelihood functions associated with (3) and (4), J refers to the number of iterations involved in the numerical maximization exercise under (3),[5] and the subscript FCM stands for fixed-coefficient model.

[4]For theory and references, see Dufour (1997).

[5]We will show in what follows that taking J explicitly into consideration has an important bearing on empirical practice.

It would be wrong, however, to compare this test statistic with tabulated values from a standard χ^2 table. Andrews (2000, 2001) has shown that the limiting null hypothesis distribution of statistics of the QLR(J)-type is not χ^2, since the parameters

$$\overline{\omega} = \left\{\sigma_{v_1}^2, \sigma_{v_2}^2, c_3, c_4\right\} \tag{6}$$

are not identified under the null model (4).

To conduct statistical testing in this context, we therefore resort to maximized Monte Carlo (MMC) tests developed by Dufour (2004),[6] which we apply to our QLR. We denote the vector of nuisance parameters that appear (and are identifiable) under the null hypothesis as:

$$\omega = \left\{\lambda, \phi_2, c_2, \sigma_\epsilon^2\right\}, \ \lambda = c_1 + \phi_1. \tag{7}$$

In Appendix A, we provide a formal exposition of the MMC test method based on any test statistic whose null distribution can be simulated and given a vector of nuisance parameters ω. Herein, we summarize the technique as it applies to our testing problem,[7] where ω is given by (7). In our test procedure, all that is needed to obtain draws from the null data-generating process is to set a value for ω; the unidentified nuisance parameters $\overline{\omega}$ simply do not intervene. Practically, we test a given TVP model as follows:

(i) We calculate the likelihood ratio statistic (5) using the likelihood values of the TVP model (3) (*the alternative model*) against its equivalent constant-coefficient model (4) (*the null model*). In the process, we save: (a) the quasi-maximum-likelihood estimate of ω imposing (4), and (b) the number of iterations the maximum-likelihood algorithm takes to converge. We denote these saved values as $\widehat{\omega}_{FCM}$ and $\widehat{J}_{TVP}$, respectively, and the observed value of (5) as $QLR_0(\widehat{J}_{TVP})$.

(ii) We generate data from the null model drawing from the normal distribution and setting ω to its estimated $\widehat{\omega}_{FCM}$ value. With this data, we re-estimate the null and the alternative models (setting the number of iterations for the TVP maximum-likelihood algorithm to $\widehat{J}_{TVP}$), and calculate the QLR statistic based on the obtained

[6]For further discussion on the Monte Carlo test method in econometrics, see, for example, Dufour and Kiviet (1996), Dufour and Khalaf (2001, 2002 a,b, 2003), and Khalaf and Kichian (2002).

[7]For proofs and references, see Dufour (2004).

likelihood values. The data generation and subsequent calculation of the QLR value are repeated in $N = 199$ replications. Thus, we obtain a 199-point distribution of generated-data QLR values. $QLR_0(\widehat{J}_{TVP})$ is compared with this distribution and a p-value is calculated based on the rank of $QLR_0(\widehat{J}_{TVP})$ relative to its simulated counterparts; see equations (2), (3), and (4) in Appendix A. This is denoted as the Monte Carlo (MC) p-value.

The MMC technique involves repeating step (ii) above, sweeping over combinations of admissible values of ω.[8] Thus, we obtain an MC p-value for each such combination. The MMC p-value is then the highest obtained MC p-value among these values.[9] The MMC test is significant at level α if the MMC p-value $\leq \alpha$. Of course, if the MC p-value obtained in step (ii) has already exceeded α (e.g., 5 per cent), there is no need to proceed with the maximization; this saves execution time. Test results are reported in section 3.

Finally, notice that, when maximizing the likelihood with simulated data, we set the number of numerical iterations to a fixed value, $\widehat{J}_{TVP}$, which corresponds to the number of iterations required to converge with the observed data.[10] The rationale is that, when the true TVPs are actually close to being constants, numerical convergence will become difficult despite the use of global maximizers. This means that the QLR statistic can sometimes be negative in practice. Nevertheless, the flexibility of the MMC test method allows us to circumvent this problem (see Dufour and Khalaf 2003 for details about QLR-type test criteria that are based on estimators at any step of the process by which the likelihood is maximized iteratively).

[8]The values of ω that we sweep over are set at $\widehat{\omega}_{FCM} \pm 5SE$; this is a fairly wide range for the parameters and, in the case of the variance, if it leads to the negative region, we truncate at the value 10^{-4}.

[9]Since the maximized p-value function is a non-differentiable step function, we use simulated annealing (a global non-gradient-based algorithm) to obtain the maxima.

[10]We thank Jean-Marie Dufour for this useful suggestion.

3. Empirical Results

We consider the data set analyzed by Pindyck (1999). The annual series for crude oil and bituminous coal extend from 1870 to 1996; for natural gas, the series extends from 1919 to 1996. Estimation is conducted on the logarithm of real prices.[11]

First, we test Pindyck's general TVP specification for each energy product. Thus, we apply the QLR statistic described previously for the null model (4) and the alternative hypothesis (3):

$$H_{A_1} : 0 < c_3 < 1, 0 < c_4 < 1. \tag{8}$$

Next, we apply the QLR tests for the specific alternatives that were selected by Pindyck (1999). We impose $c_3 = 1$ for oil, $c_4 = 1$ for coal, and omit the time-varying drift for gas. These alternatives are formally described as:

$$H_{A_2} : \quad c_3 = 1 \text{ and } 0 < c_4 < 1, \tag{9}$$

$$H_{A_3} : \quad c_4 = 1 \text{ and } 0 < c_3 < 1, \tag{10}$$

$$H_{A_4} : \quad \phi_{1t} = \phi_1, \ t = 1, ..., T \text{ and } 0 < c_4 < 1. \tag{11}$$

Table 1 reports the results of the MC and MMC test p-values. All the MC tests are applied with 199 replications and the algorithm that maximizes the p-value function (in terms of ω) is initialized at the MC point. Table 2 reports parameter estimates for significant models.

The statistical tests show that the general TVP specification is supported by the data for natural gas prices (at the 1 per cent level) and coal prices (at the 7 per cent level), but not by the data for oil prices. As for the alternatives reported by Pindyck, we find support for coal prices (at the 8 per cent level) and for gas prices (at the 1 per cent level), but not for oil prices.

[11]The data were generously provided by Pindyck. The nominal oil price series and the natural gas series for 1870 to 1973 are from Manthy (1978) and the U.S. Bureau of the Census (1975). Pindyck (1999) updated this series through 1995 using data from the U.S. Energy Information Agency and, for 1996, from the *Wall Street Journal*. The series are deflated using the U.S. wholesale price index until 1970, and the producers price index thereafter.

Table 1 - Simulation-Based Test Results

Description of tested model	Energy type	MC p-val	MMC p-val
General TVP model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_3, c_4 < 1$	Oil	0.57	-
	Coal	0.06	0.07
	Gas	0.01	0.01
Pindyck oil model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = \phi_{1t-1} + v_{1t}$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_4 < 1$	Oil	0.43	-
Pindyck coal model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = \phi_{2t-1} + v_{2t}$ $0 < c_3 < 1$	Coal	0.02	0.08
Pindyck gas model $P_t = \lambda + c_2 P_{t-1} + \phi_{2t} t + \epsilon_t$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_4 < 1$	Gas	0.01	0.01

Notes: MC p-val and MMC p-val designate Monte Carlo and maximized Monte Carlo p-values, respectively. For the MMC, parameter search spaces covered the ordinary least squares (OLS) estimate $\pm$ 5 times its standard deviation, except for the OLS variance where we used the range [0.0001,2]. 199 replications are used for each MC exercise.

Table 2 - Parameter Estimates for Significant Models

Description of model	Energy	c_1	c_2	c_3	c_4	σ^2_ϵ	$\sigma^2_{v_1}$	$\sigma^2_{v_2}$
General TVP model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_3, c_4 < 1$	Coal	0.37	0.79	0.22	0.93	#	0.0736	0.0002
	Gas	2.71	0.52	0.08	0.96	#	0.0312	0.0015
Pindyck coal model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = \phi_{2t-1} + v_{2t}$ $0 < c_3 < 1$	Coal	0.35	0.70	0.17	1.00	0.0001	0.0773	0.0002
Pindyck gas model $P_t = \lambda + c_2 P_{t-1} + \phi_{2t} t + \epsilon_t$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_4 < 1$	Gas	0.07	0.96	-	0.74	0.0434	-	0.0015

Note: # indicates a value that is less than $1x10^{-4}$.

The results indicate good overall statistical support for Pindyck's proposed class of models: there are significant TVP effects in two of the three price series examined. Our conclusions, however, differ from Pindyck's regarding which product price is best represented by TVP models. Unlike Pindyck, we find that the evolution of natural gas and coal prices is best captured by TVP effects, and not the dynamics of oil prices. Furthermore, we obtain parameter estimates (see Table 2) that are quite different from those reported by Pindyck for comparable models.[12]

The differences in these outcomes need to be elucidated because we consider Pindyck's approach of focusing on the models' forecast content and our approach of statistically testing the models' significance to be complementary. For coal and natural gas, the answer

[12]Only significant TVP model parameters are reported in Table 2.

may well reside in the sensitivity of Kalman filter estimates to filter initializations.[13] But, for crude oil prices, the difference between our and Pindyck's outcomes is striking. It might be that, although there are TVP effects in the data, they are too small to be captured statistically.

The multiplicity of significant TVP models for coal and gas prices illustrates that, while the non-adequacy of the fixed-coefficient model is settled, the question remains as to which TVP model better fits the data. One way to distinguish between the TVP models is to assess their relative forecast performance. We thus estimate the coal and natural gas TVP models for the samples ending in 1970 and 1980, respectively, and, with the state variables fixed to their end-of-sample-values, we compute dynamic forecasts out to the end of the full sample (i.e., 1996). We then calculate the mean square errors (MSE) for each of these forecast series and report them in Table 3. For comparison purposes, we also report MSE of forecasts obtained from a fixed-coefficient model and a unit root model, both estimated over samples ending in 1970 and 1980.

In the case of coal prices, the results shown in Table 3 indicate that, over the longer horizon (column MSE1971), all four reported models have similar outcomes. The generalized TVP model, however, has an MSE that is 10 times smaller than that of the corresponding fixed-coefficient model over the shorter forecast range. Similarly, in the case of natural gas prices, although the forecast performances of all four models are fairly comparable over the 25-year horizon, the specification chosen by Pindyck yields an MSE that is 30 times smaller than its fixed-coefficient counterpart over the 15-year forecast range. In contrast, the TVP model for oil has a markedly worse outcome than its corresponding fixed-coefficient model, particularly over the shorter horizon. Finally, it is interesting to note that forecast errors are generally much smaller for the coal series than for the gas series.

The computation of the MSE confirms the outcomes of the statistical tests and illustrates the usefulness of TVP-type models for forecasting coal and natural gas prices. In fact, the MSE exercise does not fully reflect the extent to which these models work better (in terms of forecasting) than their fixed-coefficient counterparts: in the forecasting

[13] A point raised by Pindyck.

Table 3 - Mean Square Errors for Selected Models

Description of forecasted model	Energy type	MSE1971	MSE1981
$P_t = \lambda + c_2 P_{t-1} + \phi_2 t + \epsilon_t$	Oil	0.378	0.086
	Coal	0.099	0.108
	gas	1.676	3.096
$P_t = \lambda + P_{t-1} + \phi_2 t + \epsilon_t$	Oil	0.206	1.241
	Coal	0.105	0.323
	Gas	1.161	1.005
General TVP model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_3, c_4 < 1$	Oil	0.446	0.441
	Coal	0.123	0.015
	Gas	1.406	2.351
Pindyck coal model $P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$ $\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$ $\phi_{2t} = \phi_{2t-1} + v_{2t}$ $0 < c_3 < 1$	Coal	0.091	0.204
Pindyck gas model $P_t = \lambda + c_2 P_{t-1} + \phi_{2t} t + \epsilon_t$ $\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$ $0 < c_4 < 1$	Gas	1.834	0.106

Notes: Mean square errors are calculated over 1971–96 and 1981–96, and denoted MSE1971 and MSE1981, respectively. The models in the first two rows are estimated by ordinary least squares.

experiment, state variables are not allowed to adjust to new information after 1970 and 1980; therefore, the MSE are actually overstated in those cases.

4. Conclusion

This paper has tested the statistical significance of Pindyck's (1999) suggested class of econometric equations that model the behaviour of long-run real energy prices. The models postulate mean-reverting prices with continuous and random changes in their level and trend, and are estimated using Kalman filtering. In such contexts, test statistics are typically non-standard and depend on nuisance parameters. Using simulation-based procedures to address this issue, we have reported results for both a standard Monte Carlo test and a maximized Monte Carlo test. Our findings lend some support to the proposed TVP class of energy models against the null hypothesis of fixed-coefficient mean-reverting equations. That is, we have found statistically significant instabilities for coal and natural gas prices, but not for crude oil prices. We have conducted out-of-sample forecasting exercises to differentiate between various significant models. This illustrates the complementarity of statistical testing and forecast analysis.

Bibliography

Amano, R. and S. Van Norden. 1995. "Exchange Rates and Oil Prices." Bank of Canada Working Paper No. 95-8.

Andrews, D.W.K. 2000. "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space." *Econometrica* 68: 399-405.

——. 2001. "Testing When a Parameter is on the Boundary of the Maintained Hypothesis." *Econometrica* 69: 683-734.

Chacra, M. 2002. "Oil Price Shocks and Retail Energy Prices in Canada." Bank of Canada Working Paper No. 2002-38.

Dufour, J.-M. 1997. "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models." *Econometrica* 65: 1365-88.

——. 2004. "Monte Carlo Tests in the Presence of Nuisance Parameters with Econometric Applications." *Journal of Econometrics*, forthcoming.

Dufour, J.-M., A. Farhat, L. Gardiol, and L. Khalaf. 1998. "Simulation-Based Finite Sample Normality Tests in Linear Regressions." *The Econometrics Journal* 1: 154-73.

Dufour, J.-M. and L. Khalaf. 2001. "Monte Carlo Test Methods in Econometrics." In *Companion to Theoretical Econometrics*, edited by B. Baltagi, 494-519. Blackwell.

——. 2002a. "Exact Tests for Contemporaneous Correlation of Disturbances in Seemingly Unrelated Regressions." *Journal of Econometrics* 106: 143-70.

——. 2002b. "Simulation-Based Finite and Large Sample Tests in Multivariate Regressions." *Journal of Econometrics* 111(2): 303-22.

——. 2003. "Finite Sample Simulation-Based Tests in Seemingly Unrelated Regressions." In *Computer-Aided Econometrics*, edited by D.E.A. Giles, Chapter 2, 11-36. Marcel Dekker.

Dufour, J.-M., L. Khalaf, J.-T. Bernard, and I. Genest. 2004. "Simulation-based Exact Tests for Heteroskedasticity and ARCH Effects."

Dufour, J.-M. and J. Kiviet. 1996. "Exact Tests for Structural Change in First-Order Dynamic Models." *Journal of Econometrics* 70: 39-68.

Hamilton, J. D. 2003. "What Is an Oil Shock?" *Journal of Econometrics* 113: 363-98.

Hansen, B. E. 1996. "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis." *Econometrica* 64: 413-30.

——. 1997. "Approximate Asymptotic *p*-values for Structural Change Tests." *Journal of Business and Economic Statistics* 15: 60-67.

Khalaf, L. and M. Kichian. 2002. "Simulation-Based Tests of Pricing-to-Market." In *Computational Methods in Decision-Making, Economics and Finance*, edited by E. Kontoghiorghes, B. Rustem, and S. Siokos, Chapter 29, 583-603. Kluwer Applied Optimization Series.

Kim, C-J., and C. R. Nelson. 1989. "The Time-Varying-Parameter Model for Modeling Changing Conditional Variance: The Case of the Lucas Hypothesis." *Journal of Business and Economic Statistics* 7(4): 433-40.

——. 1999. *State-Space Models with Regime-Switching.* MIT Press.

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*, 2nd ed. New York: Wiley.

Manthy, R. S. 1978. *Natural Resource Commodities: A Century of Statistics.* Baltimore: John Hopkins Press.

Pindyck R. S. 1999. "The Long Run Evolution of Energy Prices." *The Energy Journal* 20: 1-27.

Saphores, J.-D., L. Khalaf, and D. Pelletier. 2002. "On Jumps and ARCH Effects in Natural Resource Prices. An Application to Stumpage Prices from Pacific Northwest National Forests." *American Journal of Agriculture Economics* 84(2): 387-400.

U.S. Bureau of the Census. 1975. *Historical Statistics of the United States, Millenial Edition, Colonial Times to the Present.* Cambridge University Press.

Zellner, A. 2002. "My Experiences with Nonlinear Dynamic Models in Economics." *Studies in Nonlinear Dynamics & Econometrics* 6(2): Article 1.

http://www.bepress.com/snde/vol6/iss2/art1.

Appendix A: The Maximized Monte Carlo Test

Monte Carlo tests have recently been generalized to the nuisance parameters-dependent case by Dufour (2004). Dufour's general MC test methodology can be summarized as follows. Consider a (continuous) test statistic, S, the null distribution of which is simulable conditional on a finite set of nuisance parameters. Conforming with the notational framework of section 2, let us denote the nuisance parameter vector $\omega \in \Omega$, and let Ω_0 refer to the nuisance parameter subspace compatible with the null hypothesis H_0 under test.

Denote by S_0 the observed value of S and let S_j, $j = 1, \ldots, N$ refer to N i.i.d. random draws from the statistic's null distribution conditional on ω. Dufour's (2004) *maximized* Monte Carlo (MMC) test is defined by the critical region

$$\sup_{\omega \in \Omega_0} [\widehat{p}_N(S_0|\omega)] \leq \alpha, \tag{1}$$

where

$$\widehat{p}_N(S_0|\omega) = \frac{N\widehat{G}_N(S_0|\omega) + 1}{N+1}, \tag{2}$$

$$\widehat{G}_N(S_0|\omega) = \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty]}(S_i - S_0), \tag{3}$$

$$I_A(x) = \begin{cases} 1, & if\ x \in A, \\ 0, & if\ x \notin A. \end{cases}$$

Note that $N\widehat{G}_N(S_0|\omega)$ is the number of simulated criteria $\geq S_0$; therefore, the formula for $\widehat{p}_N(S_0|\omega)$ gives a conditional *empirical p-value*. Dufour (2004) demonstrates that the MMC test based on (1) is exact at level α:

$$\mathsf{P}_{(H_0)}\left\{\sup_{\omega \in \Omega_0} [\widehat{p}_N(S_0|\omega)] \leq \alpha\right\} \leq \alpha,$$

where $\mathsf{P}_{(H_0)}$ refers to the probability under H_0. The only conditions underlying the latter inequality are: (i) the possibility of simulating the relevant test statistic under the null

hypothesis, and (ii) $\alpha(N+1)$ is an integer.[14] No asymptotics on N or T (neither the number of replications nor the sample size) are required.

In this context, given any consistent estimate of ω that satisfies H_0 (denoted $\widehat{\omega}$), a parametric bootstrap-type critical region can be obtained as:

$$\widehat{p}_N(S_0|\widehat{\omega}) \leq \alpha. \tag{4}$$

In general, however, nothing guarantees that the level property,

$$\mathsf{P}_{(H_0)}\left[\widehat{p}_N(S_0|\omega) \leq \alpha\right] \leq \alpha,$$

holds. Under specific regularity conditions, the bootstrap p-value may be valid asymptotically in the sense that

$$\lim_{T\to\infty}\left\{\mathsf{P}[\widehat{p}_N(S_0|\widehat{\omega}) \leq \alpha] - \mathsf{P}[\widehat{p}_N(S_0|\omega_0) \leq \alpha]\right\} = 0, \tag{5}$$

where $\widehat{p}_N(S_0|\omega_0)$ is the empirical p-value that one would obtain for the "true" (unknown) nuisance parameters values. Unfortunately, in the context of the TVP test, the results of Dufour (1997, 2004) and Andrews (2000, 2001) imply that the conditions underlying (5) fail for the same reason that standard asymptotics fail. In practice, this means that, if a test rejects based on (4), this result may be spurious *even in large samples*. Yet bootstrap non-rejections are not subject to the same limitations: if the bootstrap type test is not significant, then we can be sure that the exact MMC test is not significant at level α. Indeed,

$$\widehat{p}_N(S_0|\widehat{\omega}) > \alpha \Rightarrow \sup_{\omega\,\in\,\Omega_0}\left[\widehat{p}_N(S_0|\omega)\right] > \alpha.$$

It is thus a good strategy to start the MMC sup-p-value step using a commonly used (e.g., a constrained QMLE) estimate of ω.

[14] For example, for $\alpha = 0.05$, N can be as low as 19. Although, in principle, raising N will typically increase the test power and decrease its sensitivity to the underlying randomization, the simulation results reported in Dufour and Khalaf (2001, 2002 a,b, 2003), Dufour et al. (1998), and Dufour et al. (2004) suggest that increasing N beyond 99 has only a small effect on power.

Appendix B: Kalman Filtering and the TVP Model

This appendix draws heavily on Kim and Nelson (1999), chapter 3. Consider the TVP model (3), which we rewrite for convenience in matrix notation as:

$$\begin{aligned} y_t &= H_t\beta_t + Az_t + \epsilon_t, \\ \beta_t &= F\beta_{t-1} + \eta_t, \ t = 1, ..., T, \\ \epsilon_t &\sim i.i.d.\ N(0, R), \\ \eta_t &\sim i.i.d.\ N(0, Q), \end{aligned} \tag{1}$$

where

$$\begin{bmatrix} y_t \end{bmatrix} = \begin{bmatrix} P_t \end{bmatrix} \quad \begin{bmatrix} H_t \end{bmatrix} = \begin{bmatrix} 1 & t \end{bmatrix},$$
$$\begin{bmatrix} \beta_t \end{bmatrix} = \begin{bmatrix} \phi_{1t} \\ \phi_{2t} \end{bmatrix} \quad \begin{bmatrix} z_t \end{bmatrix} = \begin{bmatrix} 1 \\ P_{t-1} \end{bmatrix},$$
$$A = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \quad \begin{bmatrix} \eta_t \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix},$$
$$F = \begin{bmatrix} c_3 & 0 \\ 0 & c_4 \end{bmatrix} \quad Q = \begin{bmatrix} \sigma^2_{v_1} & 0 \\ 0 & \sigma^2_{v_2} \end{bmatrix}.$$

The prediction equations in the Kalman filter algorithm are given by:

$$\beta_{t|t-1} = F\beta_{t-1|t-1}, \tag{2}$$
$$S_{t|t-1} = FS_{t-1|t-1}F' + Q, \tag{3}$$

where $\beta_{t|t-1}$ is the forecast value of β_t on the basis of information available through date $t-1$, and $S_{t|t-1}$ is its conditional variance. The conditional forecast error and its conditional variance can be obtained as:

$$e_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t, \tag{4}$$
$$f_{t|t-1} = H_tS_{t|t-1}H_t' + R. \tag{5}$$

These expressions can be used in the updating equations of the algorithm according to

$$\beta_{t|t} = \beta_{t|t-1} + K_te_{t|t-1}, \tag{6}$$
$$S_{t|t} = S_{t|t-1} - K_tH_tS_{t|t-1}, \tag{7}$$

where the Kalman gain term is $K_t = S_{t-1|t-1} H_t' f_{t|t-1}^{-1}$.

If, in addition to the error terms ϵ_t and η_{it}, the initial value of β is Gaussian, then the distribution of y_t conditional on information available through time $t-1$ is also Gaussian, and its log-likelihood function is:

$$lnL = -(1/2) \sum_{t=1}^{T} ln(2\pi f_{t|t-1}) - (1/2) \sum_{t=1}^{T} e_{t|t-1}' f_{t|t-1} e_{t|t-1}. \tag{8}$$

Therefore, given initial values for model parameters and state variables, the log-likelihood function can be maximized over the sample to yield maximum-likelihood parameter estimates. See Kim and Nelson (1999) for additional details.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue.***

2004

2004-4	A Structural Small Open-Economy Model for Canada	S. Murchison, A. Rennison, and Z. Zhu
2004-3	Modélisation << PAC >> du secteur extérieur de l'économie américaine	M.-A. Gosselin and R. Lalonde
2004-2	Exact Tests of Equal Forecast Accuracy with an Application to the Term Structure of Interest Rates	R. Luger
2004-1	The Effect of Adjustment Costs and Organizational Change on Productivity in Canada: Evidence from Aggregate Data	D. Leung

2003

2003-44	Common Trends and Common Cycles in Canadian Sectoral Output	F. Barillas and C. Schleicher
2003-43	Why Does Private Consumption Rise After a Government Spending Shock?	H. Bouakez and N. Rebei
2003-42	A Structural VAR Approach to the Intertemporal Model of the Current Account	T. Kano
2003-41	Anatomy of a Twin Crisis	R.H. Solomon
2003-40	Poignée de main invisible et persistance des cycles économiques : une revue de la littérature	C. Calmès
2003-39	Alternative Targeting Regimes, Transmission Lags, and the Exchange Rate Channel	J.-P. Lam
2003-38	Simple Monetary Policy Rules in an Open-Economy, Limited-Participation Model	S. Hendry, W-M. Ho, and K. Moran
2003-37	Financial Constraints and Investment: Assessing the Impact of a World Bank Loan Program on Small and Medium-Sized Enterprises in Sri Lanka	V. Aivazian, D. Mazumdar, and E. Santor
2003-36	Excess Collateral in the LVTS: How Much is Too Much?	K. McPhail and A. Vakos

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca